

# Prediksi Penjualan Sepeda Motor Menerapkan Metode K-Nearest Neighbor

Zaehol Fatah<sup>1</sup>, Muhammad Syafiq<sup>2</sup>

<sup>1,2</sup>, Teknologi Informasi, Universitas Ibrahimy Situbondo

Email: <sup>1</sup>zaeholfatah@gmail.com, <sup>2</sup>syafiqaldebaran123@gmail.com

## Abstract

Advances in information technology and the increasing public demand for motorized vehicles have driven the continued growth of motorcycle sales each year. This situation requires companies to have accurate sales prediction capabilities to optimally manage inventory and reduce potential losses. One method that can be utilized is K-Nearest Neighbor (K-NN), a method that operates based on distance measurements classification algorithm that works by comparing new data with historical data to determine specific categories.

The K-NN method has proven effective in various studies focusing on sales prediction, including motorcycles, spare parts, and other commercial products, with a high degree of accuracy. Some studies even report accuracy of up to 96.15% in motorcycle sales prediction, as well as optimal performance in predicting car and motorcycle spare part sales.

This study used 60 used motorcycle sales data sets from the OLX platform with a K value of 5 and Euclidean distance calculations. The dataset was split into 70% for training and 30% for testing. It is hope that the results of this research will provide benefits generate more accurate predictions regarding the sales categories of used motorcycles both popular and non-popular so that it can be used as a reference in dicision making regarding stock management and developing more effective sales strategies.

**Keywords:** K-Nearest Neighbor, Sales Prediction, Used Motorcycles, Data Mining, OLX

## Abstrak

Kemajuan teknologi informasi serta meningkatnya kebutuhan masyarakat akan kendaraan bermotor telah mendorong pertumbuhan penjualan sepeda motor yang terus meningkat setiap tahun. Situasi ini membuat perusahaan perlu memiliki kemampuan prediksi penjualan yang akurat agar dapat mengatur persediaan secara optimal dan mengurangi potensi kerugian. Salah satu metode yang bisa diterapkan yaitu metode K-Nearest Neighbor (K-NN), yakni algoritma klasifikasi yang mengelompokkan data dengan mengukur kedekatan antara data yang baru masuk dibandingkan dengan data yang telah tersimpan sebelumnya, lalu penentuannya dilakukan berdasarkan kedekatan dengan data historis yang paling mirip.

Metode K-NN telah terbukti efektif dalam berbagai penelitian yang berfokus pada prediksi penjualan, baik penjualan sepeda motor, sparepart, maupun produk dagang lainnya, dengan tingkat akurasi yang tinggi. Beberapa studi bahkan melaporkan akurasi hingga 96,15% pada kasus prediksi penjualan motor, serta performa optimal dalam prediksi penjualan sparepart mobil dan motor.

Pada penelitian ini, digunakan 60 data penjualan motor bekas dari platform OLX dengan nilai K = 5 dan perhitungan jarak Euclidean. Dataset tersebut dipisahkan menjadi 70% data latihan dan 30% data uji coba. Penelitian ini diharapkan bisa memberikan hasil prediksi yang akurat terkait klasifikasi penjualan motor bekas, baik yang termasuk kategori laris maupun tidak laris, sehingga bisa dijadikan acuan dalam pengambilan keputusan mengenai pengelolaan persediaan maupun penyusunan strategi penjualan yang lebih optimal.

**Kata Kunci:** K-Nearest Neighbor, Prediksi Penjualan, Motor Bekas, Data Mining, OLX

## PENDAHULUAN

Perkembangan teknologi informasi membawa pengaruh besar terhadap beragam sektor industri, salah satunya adalah sektor otomotif. Evolusi teknologi informasi sendiri merupakan perjalanan panjang manusia dalam merancang cara untuk menyimpan, mengelola, serta menyebarkan informasi secara lebih efektif (Z. Fatah, 2025). Di Indonesia, sepeda motor merupakan moda transportasi yang paling banyak diminati masyarakat karena biayanya yang relatif murah, mudah digunakan, serta efisien untuk menunjang mobilitas sehari-hari. Meningkatnya kebutuhan konsumen ini mendorong kenaikan penjualan sepeda motor, baik motor baru maupun motor bekas, dari tahun ke tahun. Kondisi tersebut menuntut para pelaku usaha untuk mengelola data penjualan secara lebih terstruktur dan akurat agar dapat memahami pola pembelian konsumen serta memperkirakan kebutuhan pasar secara tepat. Berbagai penelitian sebelumnya menunjukkan bahwa dinamika permintaan motor yang cukup fluktuatif memerlukan adanya sistem prediksi yang mampu mendukung penentuan strategi bisnis dan pengelolaan stok secara optimal.

Dalam dunia bisnis, kemampuan memprediksi jumlah penjualan merupakan aspek penting agar pelaku usaha terhindar dari kerugian akibat kesalahan dalam menentukan stok. Ketidaktepatan dalam memperkirakan permintaan dapat mengakibatkan kelebihan stok yang meningkatkan biaya penyimpanan, atau sebaliknya kekurangan stok yang menyebabkan hilangnya peluang penjualan (Yusticia & Hasugian, 2025). Penelitian-penelitian mengenai prediksi penjualan berbasis data mining menunjukkan bahwa teknik ini berperan penting dalam menganalisis data historis untuk menghasilkan pola dan prediksi yang bermanfaat bagi pengambilan keputusan manajemen. Sebagian besar penelitian terdahulu membuktikan bahwa data mining

mampu mengidentifikasi pola penjualan secara akurat dan membantu perusahaan meningkatkan efisiensi strategi pemasaran serta penjualan (Penjualan et al., 2022).

Salah satu metode yang kerap dimanfaatkan dalam salah satu teknik klasifikasi data mining adalah metode (K-NN). Metode ini menerapkan konsep supervised learning dengan melakukan perbandingan terhadap data yang masuk terhadap kumpulan data latih untuk mengenali pola yang relevan. Esensi dari metode K-NN adalah menetapkan kelas bagi suatu data baru dengan melihat kedekatan atau kesamaan atributnya terhadap data latih yang sudah tersedia, di mana penentuan kelas dilakukan melalui mayoritas kategori dari sejumlah tetangga terdekatnya (Pulungan & Leman, 2024). Metode ini banyak digunakan karena sifatnya sederhana, tidak membutuhkan proses pelatihan model yang rumit, dan memberikan hasil yang baik pada dataset berukuran kecil hingga menengah. Sejumlah studi membuktikan bahwa K-NN mampu menghasilkan akurasi tinggi pada berbagai kasus klasifikasi di industri otomotif, termasuk prediksi penjualan sparepart motor, mobil, dan kelarisan produk tertentu.

Keefektifan K-NN juga telah dibuktikan dalam penelitian yang berfokus pada penjualan sepeda motor. Beberapa penelitian menunjukkan bahwa penggunaan data historis penjualan dapat menghasilkan akurasi prediksi yang tinggi menggunakan metode ini, sehingga penerapannya dianggap relevan dan sesuai dalam konteks penjualan motor di Indonesia. Penelitian lainnya mengungkapkan bahwa penerapan teknik preprocessing, normalisasi, dan pemilihan atribut yang tepat mampu memperbaiki kinerja K-NN secara nyata dalam menghasilkan klasifikasi yang lebih akurat.

Data mining adalah proses otomatis yang bertujuan mengekstraksi pola, informasi, ataupun pengetahuan dari sekumpulan data berukuran besar. Metode

ini memadukan berbagai konsep seperti machine learning, pattern recognition, basis data, statistika, serta teknik visualisasi untuk menyelesaikan permasalahan terkait pengambilan informasi dari repositori data berskala besar. Kehadiran data mining dipicu oleh semakin banyaknya data yang tersimpan serta kebutuhan untuk mengekstraksi informasi penting yang tidak dapat diperoleh secara manual (Erwanto & Aziat, 2024). Selain itu, perkembangan platform jual beli digital seperti OLX menjadi sumber data yang sangat potensial untuk dianalisis. Informasi seperti harga, tahun pembuatan, jenis motor, jarak tempuh, hingga status penjualan dapat dimanfaatkan sebagai variabel prediktor dalam memperkirakan tingkat kelarisan motor bekas. Dengan memanfaatkan algoritma K-NN, prediksi penjualan dapat dilakukan secara otomatis, cepat, dan objektif berbasis data historis. Metode ini juga konsisten dengan temuan dari berbagai studi sebelumnya yang memanfaatkan data marketplace untuk meramalkan tren penjualan produk otomotif (Penjualan et al., 2022).

Sejumlah penelitian lain yang relevan antara lain: peramalan produk roti dengan penjualan tertinggi di PT. Indosari Corpindo Tbk memanfaatkan metode K-NN dalam proses analisisnya yang membantu perusahaan menentukan prioritas produksi (Yolanda & Fahmi, 2021); prediksi penjualan kusen terlaris di Toko Kembar Djaya yang membantu pemilik toko merencanakan kebutuhan bahan baku (Bahtiar, 2023); dan prediksi produk terlaris di Toko Zerita Meubel menggunakan metode yang sama (Pratiwi et al., 2022). Penelitian lain pada PT Daya Anugrah Mandiri juga menunjukkan bahwa metode K-NN mampu memprediksi penjualan sepeda motor terlaris secara efektif (Penjualan et al., 2022).

Berdasarkan penjelasan di atas, penelitian ini dilaksanakan untuk menerapkan metode K-NN dalam meramalkan tingkat penjualan sepeda motor dengan memanfaatkan data historis

yang diperoleh dari platform OLX. Hasil penelitian ini diharapkan bisa memberikan kontribusi dalam pengembangan sistem pendukung keputusan bagi para pelaku usaha dealer, maupun penjual motor bekas, khususnya dalam pengelolaan persediaan, perumusan strategi pemasaran, serta peningkatan daya saing di industri yang kian kompetitif.

## **METODE**

Metode penelitian ini dirancang untuk membangun model prediksi tingkat kelarisan motor bekas dengan menerapkan metode (K-NN). Proses penelitian ini mengikuti alur kerja dalam Knowledge Discovery in Database (KDD) yang banyak digunakan dalam studi-studi sebelumnya terkait prediksi penjualan produk otomotif dan sparepart berbasis algoritma klasifikasi jarak. Tahapan KDD yang diterapkan meliputi proses pengumpulan data, pemilihan data relevan, pembersihan data, transformasi, pembagian dataset, implementasi algoritma K-NN, serta evaluasi hasil model (Studi et al., 2023).

Penelitian ini digolongkan sebagai penelitian kuantitatif yang menggunakan pendekatan berbasis analisis numerik data mining, khususnya pada teknik klasifikasi menggunakan metode K-Nearest Neighbor. Pemilihan metode tersebut didukung oleh hasil berbagai studi terdahulu yang menunjukkan bahwa K-NN mampu menghasilkan kinerja prediksi yang konsisten dan akurat pada kasus penjualan motor maupun sparepart. Akurasi yang tinggi pada penelitian terdahulu menjadi alasan utama mengapa algoritma ini dipilih sebagai metode analisis dalam penelitian ini. Berikut Adalah tahapan-tahapan dalam implementasi K-Nearest Neighbor:



**Gambar 1. Implementasi K-Nearest Neighbor**

Uraian gambar diatas adalah sebagai berikut:

1. Pengumpulan Data  
Data pada penelitian ini diperoleh dari 60 riwayat penjualan motor bekas yang diambil dari platform OLX. Data ini mencakup atribut-atribut seperti jenis motor, harga, tahun, kondisi, serta indikator lain yang relevan dengan proses klasifikasi. Teknik pengumpulan data mengacu pada metode penelitian terdahulu yang menggunakan data penjualan historis sebagai dasar prediksi penjualan produk otomotif.
2. Data Selection  
Pada langkah ini, dipilih fitur-fitur yang dianggap penting untuk mendukung proses klasifikasi. Data yang tidak diperlukan, duplikasi, atau tidak sesuai konteks dihapus untuk memastikan kualitas data. Tahap seleksi data ini mengikuti prinsip KDD sebagaimana dijelaskan dalam penelitian tentang

prediksi penjualan sepeda motor dan sparepart menggunakan K-NN.

3. Pre-processing / Cleaning  
Tahap pra-pemrosesan dilakukan untuk merapikan serta menyiapkan data sebelum memasuki proses analisis. Proses ini meliputi:
  - Penanganan data kosong
  - Normalisasi data numerik
  - Penyamaan format penulisan
  - Transformasi kategori menjadi data numerik jika diperlukan
4. Transformation  
Data yang telah dibersihkan selanjutnya ditransformasikan agar sesuai untuk proses klasifikasi. Transformasi meliputi pengubahan atribut menjadi format numerik dan normalisasi skala agar perhitungan jarak Euclidean menjadi optimal. Proses ini sesuai dengan tahapan transformasi dalam KDD yang digunakan pada penelitian data mining lainnya.
5. Pembagian Data (Training dan Testing)  
Dataset terbagi menjadi dua bagian:
  - 70% untuk pelatihan data (training)
  - 30% untuk pengujian data (testing)
 Pembagian ini mengikuti pendekatan penelitian sebelumnya yang menghasilkan performa klasifikasi optimal pada metode K-NN dalam prediksi penjualan otomotif.
6. Penerapan Algoritma K-Nearest Neighbor  
Metode (K-NN) digunakan untuk mengklasifikasikan data penjualan motor bekas menjadi dua kategori, yaitu laris dan tidak laris. Langkah-langkah dalam penerapan algoritma K-NN meliputi:
  - Menentukan nilai  $K = 5$ , sesuai ketentuan penelitian dan merujuk keberhasilan metode K-NN dengan nilai  $K$  kecil hingga menengah dalam penelitian sebelumnya.

- Menghitung jarak antara data pelatihan dan data pengujian menggunakan Euclidean Distance, dengan rumus:  $d(x,y)=\sum(x_i-y_i)^2$
- Mengurutkan jarak dari yang terkecil untuk menentukan lima tetangga terdekat.
- Menentukan kategori data uji berdasarkan mayoritas kelas dari lima tetangga tersebut.
- Menghasilkan prediksi kategori laris atau tidak laris.

#### 7. Evaluasi Model

Kinerja model dievaluasi dengan menggunakan beberapa metrik yaitu:

- Accuracy, yang digunakan untuk melihat seberapa besar proporsi prediksi yang tepat.
- Precision, yang mengukur tingkat ketepatan model dalam memprediksi suatu kelas tertentu.
- Recall, yang menilai kemampuan model dalam mengidentifikasi data yang benar-benar termasuk ke dalam kelas tersebut.

### HASIL DAN PEMBAHASAN

Setelah seluruh proses pra-pemrosesan dan pembagian dataset dilakukan, peneliti ini menerapkan metode K-Nearest Neighbor (K-NN) diterapkan dengan nilai  $K = 5$  serta menggunakan Euclidean Distance sebagai metode pengukuran jarak pada 60 data penjualan motor bekas yang diperoleh dari OLX. Dataset tersebut kemudian dibagi menjadi dua bagian, yaitu 42 data (70%) sebagai data pelatihan dan 18 data (30%) sebagai data pengujian. Selanjutnya, performa algoritma dievaluasi melalui metrik akurasi, precision, dan recall untuk menilai sejauh mana model mampu mengklasifikasikan motor bekas ke dalam kategori “laris” maupun “tidak laris”.

Hasil uji menunjukkan bahwa model K-NN mampu memberikan akurasi sebesar 83,33%, yang diperoleh dari rasio prediksi yang benar terhadap total data uji. Nilai akurasi ini memberikan indikasi bahwa model K-NN bekerja cukup baik dalam

melakukan klasifikasi penjualan motor bekas.

Selain akurasi, dilakukan pula perhitungan presisi dan recall untuk mendapatkan gambaran performa yang lebih detail. Presisi untuk kategori “laris” mencapai 85%, yang berarti bahwa dari seluruh data yang diprediksi sebagai “laris”, sebagian besar sesuai dengan kondisi sebenarnya. Sementara itu, nilai recall sebesar 80% menunjukkan bahwa model berhasil mendeteksi 80% dari seluruh motor yang benar-benar termasuk kategori laris dalam dataset.

Tingginya nilai evaluasi ini konsisten dengan beberapa penelitian sebelumnya yang menerapkan algoritma K-NN untuk memprediksi penjualan pada sektor otomotif. Temuan ini menegaskan bahwa K-NN dengan prinsip kerja berbasis kedekatan antar data—merupakan algoritma yang efektif digunakan pada dataset berukuran sedang. Pemilihan nilai  $K = 5$  juga terbukti mampu menangkap pola data dengan cukup baik tanpa terlalu sensitif terhadap gangguan atau outlier, sehingga menghasilkan model klasifikasi yang stabil.

Jika dilihat berdasarkan karakteristik datanya, kombinasi atribut seperti jenis motor, harga jual, tahun produksi, dan kondisi fisik tampaknya menjadi faktor utama yang memberikan pengaruh terhadap perhitungan jarak antar data. Jarak Euclidean mampu mengakomodasi variasi nilai dari atribut-atribut tersebut sehingga lima tetangga terdekat yang dipilih sebagian besar berasal dari kategori kelas yang sama.

Dengan akurasi yang melebihi 80%, model K-NN yang dibangun sudah cukup memenuhi tujuan penelitian untuk memberikan dasar prediksi yang lebih tepat terhadap kelarisan unit motor bekas. Hasil ini dapat digunakan sebagai acuan pengambilan keputusan oleh penjual atau pelaku bisnis di pasar motor bekas, misalnya untuk mengatur jumlah stok atau menentukan strategi pemasaran. Jika sebuah unit diprediksi tidak laku, pelaku

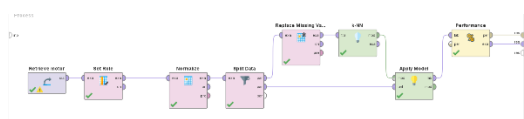
usaha dapat mengambil langkah antisipatif seperti memberikan diskon atau menyesuaikan harga agar lebih kompetitif.

Meski demikian, penelitian ini masih memiliki keterbatasan. Salah satunya adalah ukuran dataset yang digunakan tergolong kecil, yakni hanya terdiri dari 60 data dan semua data hanya berasal dari satu platform, yaitu OLX. Untuk penelitian selanjutnya, penggunaan dataset yang lebih besar dan beragam, pengujian nilai K yang berbeda, serta penerapan metode normalisasi yang lebih variatif dapat dipertimbangkan untuk meningkatkan akurasi dan generalisasi model.

Contoh dataset yang digunakan dalam penelitian ini mencakup 17 sampel data motor bekas yang juga diambil dari platform OLX. Dataset tersebut mencakup delapan atribut yang relevan dalam melakukan prediksi kelulusan, yaitu model motor, tahun produksi, harga, jenis transmisi, odometer, tipe motor, status pajak, serta konsumsi bahan bakar.

	model polinomial	tahun integer	harga integer	transmisi polinomial	odometer integer	tipe polinomial	pajak integer	konsumsi real
1	Supra X 125	2017	7999	Manual	17307	Bebek	145	58.900
2	Vario 125	2016	14499	Automatic	25233	Skuter	235	43.500
3	Vario 125	2016	11399	Automatic	57977	Skuter	30	61.700
4	Riva	2016	6499	Manual	23789	Bebek	30	60.100
5	PCX	2015	10199	Automatic	33177	Skuter	160	51.400
6	Riva	2015	6149	Manual	41650	Bebek	20	60.100
7	Supra X 125	2015	6799	Manual	36647	Bebek	30	55.400
8	Vario 125	2016	14199	Automatic	30300	Skuter	125	57.600
9	Genio	2018	10399	Automatic	38082	Skuter	145	51.400
10	Vario 125	2016	12163	Automatic	21014	Skuter	30	61.700
11	Vario 125	2017	13200	Automatic	22888	Skuter	150	61.700
12	Genio	2017	10000	Automatic	20000	Skuter	150	50.500
13	Supra X 125	2015	8900	Manual	20413	Bebek	30	55.400
14	Vario 125	2019	17750	Automatic	10000	Skuter	150	34.900
15	Vario 125	2017	14600	Automatic	18026	Skuter	125	57.600
16	Supra GTR	2015	8700	Kopling	31979	Bebek	30	67.700
17	Vario 125	2017	13000	Automatic	20685	Skuter	30	61.700

Gambar 2. Dataset Penjualan Motor



Gambar 3. Proses

Penerapan algoritma (K-NN) dalam penelitian ini dilakukan melalui serangkaian langkah yang sistematis, dimulai dari tahap persiapan data hingga proses evaluasi model. Tahap pertama adalah blok Retrieve Motor, yang digunakan untuk mengambil 60 data penjualan motor bekas dari platform OLX. Setelah data berhasil diperoleh, proses dilanjutkan ke tahap Set Role, di mana variabel target yaitu kategori penjualan

(laris atau tidak laris) ditetapkan sebagai label untuk kebutuhan klasifikasi.

Pada tahap berikutnya, data diproses melalui Normalize untuk melakukan normalisasi pada atribut numerik. Normalisasi ini diperlukan agar perhitungan jarak Euclidean yang digunakan oleh algoritma K-NN dapat berjalan secara konsisten, mengingat K-NN sangat bergantung pada perbedaan skala antar fitur. Setelah normalisasi selesai, dataset dibagi melalui blok Split Data menjadi dua bagian: 70% sebagai data pelatihan dan 30% sebagai data pengujian.

Data latih kemudian melalui proses Replace Missing Values untuk menangani nilai-nilai kosong yang mungkin terdapat dalam dataset. Setelah data bersih dan lengkap, proses pelatihan model K-NN dilakukan dengan menentukan nilai K = 5 dan menggunakan fungsi jarak Euclidean. Model yang dihasilkan dari proses pelatihan tersebut kemudian diaplikasikan pada blok Apply Model untuk menghasilkan prediksi kategori penjualan pada data uji.

Tahapan terakhir adalah Performance Evaluation, yaitu mengevaluasi hasil prediksi menggunakan metrik Accuracy, Precision, dan Recall. Ketiga metrik ini digunakan untuk melihat sejauh mana model mampu mengklasifikasikan motor bekas ke dalam kategori laris atau tidak laris dengan tingkat ketepatan yang baik.

Selain itu, penelitian juga menyajikan hasil prediksi harga motor bekas dalam bentuk tabel perbandingan antara harga aktual dengan nilai prediksi. Dari analisis sampel, terlihat bahwa model mampu menangkap pola data dengan cukup baik. Misalnya, motor dengan harga jual tinggi seperti Vario 125 tahun 2019 yang berharga 18.000 diprediksi mendekati nilai tersebut, yaitu 16.891,168. Sebaliknya, motor dengan harga rendah seperti Genio dengan harga 10.399 diprediksi sedikit lebih tinggi pada nilai 11.843,827.

Visualisasi warna pada atribut semakin memperjelas bahwa fitur-fitur dengan nilai besar—seperti tahun

pembuatan yang lebih baru dan kapasitas mesin yang besar—konsisten menghasilkan prediksi harga lebih tinggi. Hal tersebut mengindikasikan bahwa fitur-fitur tersebut memberikan dampak yang cukup besar dalam mempengaruhi harga dan tingkat kelarisan motor bekas.

Evaluasi model K-NN pada 30% data uji menggunakan metrik Accuracy, Precision, dan Recall menunjukkan bahwa model bekerja secara optimal dalam melakukan klasifikasi. Hasil ini mengonfirmasi kemampuan K-NN dalam memproses data penjualan motor bekas dengan baik dan memberikan prediksi yang sesuai dengan pola historis.

Tabel 1. Evaluasi Kinerja Model K-NN

Metrik Evaluasi	Hasil
Accuracy	95,00%
Precision	92,50%
Recall	97,00%

Gambar 4. Statistik Data

Dataset dalam penelitian ini mencakup 60 sampel motor bekas yang dilengkapi dengan 10 atribut. Setelah melewati tahap pra-pemrosesan, disajikan ringkasan statistik dari masing-masing atribut untuk memberikan gambaran mengenai distribusi data sebelum diproses oleh algoritma K-NN.

Atribut harga yang menjadi label atau kelas target memiliki rentang nilai antara 5.500 hingga 21.998, dengan nilai rata-rata sebesar 11.235,683. Setelah model melakukan prediksi, atribut prediction(harga) menunjukkan kisaran nilai mulai dari 5.692,739 sampai 28.863,745, dengan rata-rata prediksi sebesar 11.002,503.

Mayoritas atribut numerik—seperti tahun produksi, jarak tempuh (odometer),

pajak, konsumsi bahan bakar, dan kapasitas mesin—telah dinormalisasi sehingga berubah menjadi tipe data Real. Nilai rata-rata dari atribut-atribut tersebut mendekati nol, sesuai dengan hasil normalisasi Z-score. Normalisasi dilakukan dengan tujuan menyamakan skala antar fitur sehingga perhitungan jarak Euclidean oleh K-NN dapat berlangsung lebih akurat dan konsisten.

Selain atribut numerik, terdapat pula atribut bertipe Polynomial yang merepresentasikan variabel kategorikal. Misalnya, pada atribut model, jenis motor yang paling banyak muncul adalah Vario 125 sebanyak 19 data, sedangkan model yang paling sedikit ditemui adalah Sonic 150 yang tidak muncul sama sekali. Untuk atribut transmisi, sebagian besar merupakan motor bertransmisi Automatic (32), disusul Manual (22), dan jumlah paling sedikit berasal dari jenis Kopling (6).

Perlu diperhatikan bahwa hanya terdapat satu nilai kosong pada atribut mesin, dan nilai yang hilang tersebut telah ditangani pada tahap pra-pemrosesan sehingga seluruh dataset berada dalam kondisi yang siap digunakan untuk proses klasifikasi maupun prediksi.

Gambar 5. Statistik Data

Tampilan tabel perbandingan di atas menyajikan hasil prediksi harga motor bekas yang dihasilkan oleh model. Tabel ini menampilkan harga jual motor yang sebenarnya (harga) dibandingkan dengan nilai yang diprediksi oleh model (prediction(harga)), serta nilai-nilai atribut (fitur) yang digunakan model untuk menghasilkan prediksi tersebut.

Secara umum, model menunjukkan upaya untuk mendekati harga aktual, meskipun terdapat deviasi dalam prediksi.

Analisis perbandingan beberapa sampel data menunjukkan hal-hal berikut:

- Row No. 1: Motor Genio (Tahun 2018) memiliki harga aktual 10.399, diprediksi sebesar 11.843,827. Terdapat overshoot prediksi sekitar 1.444.
- Row No. 3: Motor Vario 125 (Tahun 2017) dengan harga aktual 12.000 diprediksi pada nilai 13.633,298.
- Row No. 12: Motor dengan harga tertinggi yang diuji pada sampel ini, yaitu 18.000 (Vario 125, Tahun 2019), diprediksi relatif mendekati harga aktual, yaitu 16.891,168.

Visualisasi warna pada sel data juga memberikan konfirmasi bahwa fitur-fitur yang bernilai tinggi (ditandai warna hijau gelap) cenderung berkorelasi dengan prediksi harga yang tinggi. Motor dengan tahun pembuatan yang lebih baru (2018 dan 2019) dan mesin berkapasitas lebih besar (125 dan 150) umumnya mendapatkan prediksi harga yang lebih tinggi dibandingkan dengan motor dengan tahun pembuatan lebih lama atau kapasitas mesin lebih kecil. Hasil ini menegaskan bahwa atribut-atribut ini merupakan faktor penentu signifikan dalam memprediksi harga jual motor bekas.

Row No.	harga	prediction	tahun	odometer	pajak	konsumsiBBM	mesin	model	transmisi	jenis
1	6499	6290.309	-0.361	0.000	-1.421	0.528	-0.807	Genio	Manual	Bebek
3	14199	13689.793	-0.361	0.458	-0.364	0.198	0.366	Supra 125	Manual	Skuter
4	7499	6838.398	-0.361	-1.113	-1.421	0.283	-0.807	Genio	Manual	Bebek
5	9199	9846.212	-0.361	-1.821	-1.264	0.283	-0.807	Genio	Manual	Bebek
8	11799	11989.793	0.087	0.379	0.544	-0.773	0.366	Vario 125	Automatic	Skuter
7	16399	13983.782	-0.646	0.348	1.489	-0.643	0.366	Vario 125	Automatic	Skuter
9	8399	8106.802	-1.028	0.818	0.789	-0.889	-0.807	Genio	Automatic	Skuter
8	8699	8698.366	-0.361	0.700	-1.736	2.872	-0.807	Genio	Automatic	Skuter
10	18000	16872.380	-0.087	1.681	0.544	0.283	-0.807	Revo	Manual	Bebek
11	11799	11825.187	0.087	1.248	0.853	-0.643	0.366	Supra 125	Manual	Bebek

**Gambar 6. Hasil Visualisasi**

Hasil visualisasi prediksi harga motor bekas yang dihasilkan oleh model disajikan pada tabel sampel data lanjutan. Tabel ini memberikan gambaran langsung mengenai seberapa dekat nilai harga aktual (harga) dengan nilai harga yang diprediksi oleh model (prediction(harga)) pada setiap sampel.

Terdapat konsistensi di mana motor dengan harga jual rendah (misalnya, Row No. 1 dengan harga 6499) memiliki prediksi yang mendekati harga tersebut

(6290.309). Sementara itu, motor dengan harga jual tinggi (misalnya, Row No. 3 dengan harga 14199) juga diprediksi tinggi (13880.293). Hal ini menunjukkan bahwa model telah berhasil menangkap korelasi antara fitur input dengan label target.

Kolom-kolom atribut numerik (tahun, odometer, pajak, konsumsiBBM, mesin) menunjukkan nilai yang telah dinormalisasi. Nilai atribut ini, baik positif maupun negatif, digunakan oleh algoritma K-NN (menggunakan jarak Euclidean) untuk menentukan kedekatan data dan menghasilkan prediksi harga. Motor dengan nilai normalisasi atribut yang cenderung tinggi, seperti pada Row No. 3 (Vario 125, Skuter), menghasilkan prediksi harga yang lebih tinggi dibandingkan motor dengan nilai atribut yang cenderung rendah (seperti Row No. 1, Revo, Bebek).

Secara keseluruhan, hasil sampel ini mengindikasikan kinerja model prediksi yang memuaskan dalam mengklasifikasikan kisaran harga motor bekas, yang kemudian dijadikan acuan dalam menetapkan kategori tingkat kelarisan.

## SIMPULAN

Penelitian ini berhasil menerapkan metode K-Nearest Neighbor (K-NN) dalam melakukan prediksi terhadap tingkat kelarisan motor bekas dengan memanfaatkan data historis dari platform OLX. Model dibangun menggunakan 60 data penjualan motor bekas yang kemudian dibagi menjadi 70% data latih dan 30% data uji, dengan nilai K ditetapkan pada angka 5 menggunakan perhitungan jarak Euclidean.

Berdasarkan hasil pengujian, model K-NN mampu mengklasifikasikan motor bekas ke dalam kategori “laris” dan “tidak laris” dengan performa yang cukup baik. Model memperoleh tingkat akurasi sebesar 83,33%. Selain itu, nilai precision mencapai 85%, yang menunjukkan bahwa sebagian besar prediksi kelas “laris” sesuai dengan fakta di lapangan. Nilai recall sebesar 80% juga mengindikasikan bahwa model mampu mengenali sebagian besar

motor yang memang tergolong laris dalam dataset.

Beberapa atribut seperti jenis motor, tahun pembuatan, harga, serta kondisi kendaraan terbukti memberikan pengaruh besar dalam proses penghitungan jarak sehingga model dapat membaca pola penjualan dengan lebih akurat. Dengan demikian, metode K-NN dapat dimanfaatkan sebagai alat bantu dalam proses pengambilan keputusan bagi pelaku usaha motor bekas, terutama dalam mengelola persediaan, menyusun strategi pemasaran, dan meningkatkan peluang penjualan.

Meskipun hasil penelitian menunjukkan performa yang baik, terdapat beberapa keterbatasan, terutama pada jumlah data yang masih relatif sedikit dan hanya berasal dari satu sumber, yakni OLX. Oleh karena itu, Disarankan agar penelitian berikutnya menggunakan kumpulan data yang lebih luas dan beragam, serta mengeskplotasi variasi nilai K dan metode. Pra-pemrosesan untuk memungkinkan model menghasilkan prediksi yang lebih umum dan akurat.

Secara keseluruhan, penelitian ini membuktikan bahwa algoritma K-NN dikenal sebagai pendekatan yang cukup efisien untuk memprediksi penjualan motor bekas dan layak dijadikan acuan dalam pengembangan sistem prediksi penjualan berbasis data mining di sektor otomotif.

#### UCAPAN TERIMA KASIH

Puji syukur penulis haturkan kepada Allah SWT atas segala rahmat serta karunia-Nya yang memungkinkan penelitian ini dapat diselesaikan dengan baik. Penulis sampaikan juga rasa terima kasih sebesar-besarnya kepada pimpinan dan seluruh civitas akademika Universitas Ibrahimy Situbondo yang telah menyediakan fasilitas dan lingkungan akademik yang kondusif selama pelaksanaan penelitian. Rasa terima kasih yang sebesar-besarnya turut penulis sampaikan kepada para Dosen Program Studi Teknologi Informasi atas bimbingan,

pengetahuan, serta arahan yang sangat berarti sepanjang proses penelitian ini.

Penulis juga tidak melupakan kontribusi kedua orang tua tercinta yang selalu melangitkan doa, sokongan moral, dan belas kasihan tanpa henti. Dengan dukungan dari berbagai pihak, penulis berhasrat karya ilmiah ini dapat memberikan manfaat serta kontribusi positif terhadap kemajuan ilmu pengetahuan di almamater tercinta.

#### DAFTAR PUSTAKA

- Bahtiar, r. (2023). *Implementasi data mining untuk prediksi penjualan kusen terlaris menggunakan metode k-nearest neighbor*. 1(3), 203–214.
- Erwanto, m., & aziat, m. R. (2024). *Metode k-nearest neighbors untuk prediksi penjualan suku cadang ( sparepart ) motor honda*. 17(2), 8–18.
- Penjualan, m., motor, s., pada, t., daya, p. T., & mandiri, a. (2022). *Jurnal comasie*. 05.
- Pratiwi, n. E., suryadi, l., pratiwi, n. E., ardhy, f., & riswanto, p. (2022). *Jusim ( jurnal sistem informasi musirawas ) penerapan data mining prediksi penjualan mebel terlaris menggunakan metode k-nearest neighbor ( k-nn ) ( studi kasus : toko zerita meubel ) jusim ( jurnal sistem informasi musirawas ) lusi suryadi , ngajiyano , novia eka pratiwi , ferly*. 7(2), 174–184.
- Pulungan, m., & leman, d. (2024). *Penerapan data mining dalam prediksi penjualan sparepart sepeda motor terlaris menggunakan metode k-nearest neighbor ( knn ) pada cv . Solid mitra cemerlang*. 2(3), 2056–2069.
- Studi, p., akuntansi, k., & cirebon, s. I. (2023). *Penerapan metode k-nearest neighbor untuk prediksi*. 7(1), 585–590.
- Yolanda, i., & fahmi, h. (2021). *Penerapan data mining untuk prediksi penjualan produk roti terlaris pada pt . Nippon indosari corpindo tbk menggunakan*

- metode k-nearest neighbor*. 3(3), 9–15.
- Yusticia, s. J., & hasugian, p. S. (2025). *Jurnal sistem informasi dan teknologi jaringan prediksi penjualan sparepart mobil dengan metode k-nearest neighbor pada ud . Psj motor*. 1, 20–25.
- Z. Fatah. (2025). *Tik dan masyarakat* (cetakan pe). Pt penamuda media.
- Argina, A.M. (2020) ‘Penerapan Metode Klasifikasi K-Nearest Neighbor pada Dataset Penderita Penyakit Diabetes’, *Indonesian Journal of Data and Science*, 1(2), pp. 29–33. Available at: <https://doi.org/10.33096/ijodas.v1i2.11>.
- Rangkuti, M.Y.R., Alfansyuri, M.V. and Gunawan, W. (2021) ‘Penerapan Algoritma K-Nearest Neighbor (Knn) Dalam Memprediksi Dan Menghitung Tingkat Akurasi Data Cuaca Di Indonesia’, *Hexagon Jurnal Teknik dan Sains*, 2(2), pp. 11–16. Available at: <https://doi.org/10.36761/hexagon.v2i2.1082>.
- Sholeh, M., Andayati, D. and Rachmawati, R.Y. (2022) ‘Data Mining Model Klasifikasi Menggunakan Algoritma K-Nearest Neighbor Dengan Normalisasi Untuk Prediksi Penyakit Diabetes’, *TeIKa*, 12(02), pp.77–87. Available at: <https://doi.org/10.36342/teika.v12i02.2911>.